

Dichotomization invariant log-mean linear parameterization for discrete graphical models of marginal independence

Alberto Roverato
University of Bologna
`alberto.roverato@unibo.it`

February 20, 2013

Abstract

We extend the log-mean linear parameterization introduced by Roverato et al. (2013) for binary data to discrete variables with arbitrary number of levels, and show that also in this case it can be used to parameterize bi-directed graph models. Furthermore, we show that the log-mean linear parameterization allows one to simultaneously represent marginal independencies among variables and marginal independencies that only appear when certain levels are collapsed into a single one. We illustrate the application of this property by means of an example based on genetic association studies involving single-nucleotide polymorphisms. More generally, this feature provides a natural way to reduce the parameter count, while preserving the independence structure, by means of substantive constraints that give additional insight into the association structure of the variables.

Keywords: Contingency table; Graphical Markov model; Marginal independence; Parsimonious model; Single-nucleotide polymorphism.

1 Introduction

Graphical models of marginal independence use a graph where every vertex is associated with a variable and missing edges encode marginal independence relationships according to a given Markov property; see Pearl and Wermuth (1994); Kauermann (1996); Banerjee and Richardson (2003); Richardson (2003). These models were introduced by Cox and Wermuth (1993, 1996) as *covariance graphs*, with dashed lines to represent edges in the graph. More recently, lines with two arrowheads are often used in place of dashed edges and, accordingly, these models are also referred to as *bi-directed graph models*. Graphical models of marginal independence have appeared in several applied contexts as described in Drton and Richardson (2008) and references therein. Their application is typically suggested when the observed variables are jointly affected by unobserved variables (see, among others, Richardson, 2003; Maathuis et al., 2009; Colombo et al., 2012) and, furthermore, they are a

special case of *acyclic directed mixed graphs* (Richardson, 2003) and can be regarded as the building blocks of *regression graph models* (Drton, 2009; Wermuth and Sadeghi, 2012).

The probability distribution of a set of discrete variables is characterized by the associated probability table, but defining a suitable parameterization for bi-directed graph models is not straightforward; see Drton and Richardson (2008); Drton (2009); Lupparelli et al. (2009); Roverato et al. (2013). A basic requirement for the flexible implementation of marginal constraints is that interaction terms involving a subset of variables satisfy *upward compatibility*, that is they should reflect a property of the corresponding marginal distribution; see Drton and Richardson (2008) for details and extensive references. Upward compatibility means invariance with respect to marginalization but, for discrete variables with arbitrary number of levels, a stronger invariance property may be required. As shown in the example below, there are situations where the research question involves different collapsed versions of a same variable. Collapsing two or more levels of a discrete variables into a single level can be regarded as a special kind of marginalization and invariance with respect to this operation is an useful feature for a parameterization.

Example 1.1 (*Genetic association analysis*). Genetic association studies aim at identifying genetic factors associated with a certain phenotype, such as a disease; see Balding (2006) for a review of statistical approaches to population association studies. Single-Nucleotide Polymorphisms (SNPs) are the most common form of variation in the human genome (see Hirschhorn and Daly, 2005). A SNP is a change in one nucleotide at a given genomic position. Commonly, SNPs are diallelic with two of the four bases *A* (adenine), *C* (cytosine), *T* (thymine) and *G* (guanine) occurring at the considered locus where it is possible that a *wild* allele, *W*, is substituted by a *mutant* allele, *M*. Hence, a SNP has three possible genotypes, *WW*, *WM* and *MM*, and can be represented as a three-level discrete variable. The latter representation makes it possible to identify the *codominant genotype effect* of the SNP on the phenotype, but the relevant phenotype may also be associated with alternative representations of the SNP. In the *dominant genotype model*, heterozygote individuals are expected to have the same phenotype as *MM* homozygote individual so that the levels *WM* and *WM* are collapsed into a single one to give *WM + MM* vs. *WW*. Conversely, in the *recessive genotype model* the three levels are dichotomized as *WM + WW* vs. *MM*. In general, none of these three genetic models for a specific SNP is favored a priori, and the research question also concerns the identification of the most appropriate representation of a SNP. Clearly, there is a loss of efficiency in fitting a different statistical model for every possible genetic model, and this is especially true when more SNPs are simultaneously considered.

In this paper we extend the *Log-Mean Linear (LML)* parameterization introduced by Roverato et al. (2013) for binary data to discrete variables with arbitrary number of levels and show that also in this case it can be used to parameterize bi-directed graph models. Furthermore, we show that the LML parameterization satisfies a stronger version of up-

ward compatibility that we call *dichotomization invariance*. Every LML parameter can be uniquely associated with a cell either of the cross-classified table or of a marginal table and it is invariant with respect to both marginalization and collapsing operations that does not involve such a cell. In this way, the LML parameterization allows one to simultaneously represent marginal independencies among variables and marginal independencies that only appear when certain levels are collapsed into a single one. This feature is useful in several applied contexts, but it also provides a natural way to reduce the parameter count, while preserving the independence structure, by means of substantive constraints that give additional insight into the dependence structure of variables. Being able to implement additional substantive constraints so as to specify parsimonious submodels is a key issue in graphical modelling and, more generally, in multivariate analysis (see, for instance, Højsgaard and Lauritzen, 2008) but it is especially relevant in marginal modelling because the number of parameters in a bi-directed graph model can be relatively large even for sparse graphs (Richardson, 2009; Evans and Richardson, 2012).

This paper is organized as follows. Sections 2 and 3 provide a review of the theory of discrete bi-directed graph models and of the associated parameterizations as required for this paper. Section 4 contains the extension of the LML parameterization to the general discrete case, whereas in Section 5 we introduce the binary expansion operation and state the connected set Markov property for B -expanded graphs. Finally, Section 6 contains a brief discussion.

2 Bi-directed graph models

Let $Y_V = (Y_v)_{v \in V}$ be a random vector with entries indexed by a finite set V . In graphical models of marginal independence every variable Y_v , with $v \in V$, is associated with a vertex of a *bi-directed graph* $\mathcal{G} = (V, E)$. The edge set E is a collection of unordered, distinct, pairs of vertices and every edge $\{i, j\} \in E$ is represented as a line with two arrowheads, $i \leftrightarrow j$. A graph is *complete* if every pair of vertices is joined by an edge. A subset $\emptyset \neq U \subseteq V$ induces a *subgraph* $\mathcal{G}_U = (U, E_U)$ where $E_U = E \cap (U \times U)$. If \mathcal{G}_U is disconnected we say that U is a *disconnected set* of \mathcal{G} and denote by C_1, \dots, C_r its inclusion maximal connected sets that we call the *connected components* of U ; see Richardson (2003) for details. Recall that $U = C_1 \dot{\cup} \dots \dot{\cup} C_r$ where the symbol $\dot{\cup}$ denotes a union of disjoint sets.

A bi-directed graph model is the family of probability distributions for Y_V that satisfy a given Markov property with respect to a bi-directed graph $\mathcal{G} = (V, E)$. The distribution of Y_V is said to satisfy the *pairwise Markov property* with respect to \mathcal{G} if for every $\{i, j\} \notin E$, with $i \neq j$, it holds that Y_i is independent of Y_j ; in symbols $Y_i \perp\!\!\!\perp Y_j$. The distribution of Y_V satisfies the *global Markov property*, also called the *connected set Markov property* by Richardson (2003), if for every disconnected set U of \mathcal{G} , the subvectors corresponding to its connected components Y_{C_1}, \dots, Y_{C_r} are mutually independent; $Y_{C_1} \perp\!\!\!\perp \dots \perp\!\!\!\perp Y_{C_r}$. We remark that the connected set Markov property implies the pairwise Markov property, whereas the

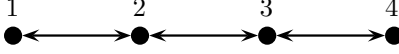


Figure 1: Bi-directed 4-chain.

converse is not true in general, even for strictly positive distributions.

Example 2.1 (*bi-directed 4-chain*). In the bi-directed graph of Figure 1 the disconnected sets are $\{1, 3\}$, $\{1, 4\}$, $\{2, 4\}$, $\{1, 2, 4\}$ and $\{1, 3, 4\}$. Under the connected set Markov property, the sets $\{1, 2, 4\}$ and $\{1, 3, 4\}$ encode the independencies $X_{\{1,2\}} \perp\!\!\!\perp X_4$ and $X_1 \perp\!\!\!\perp X_{\{3,4\}}$, respectively, and these imply the independencies encoded by each of the remaining disconnected sets.

3 Parametrizations of discrete bi-directed graph models

3.1 The Möbius and LML parameters for the binary case

In order to highlight that a variable is binary we denote it by X and, without loss of generality, we assume that $X \in \{0, 1\}$ so that $X_V = (X_v)_{v \in V}$ is a multivariate Bernoulli random vector taking values in $\{0, 1\}^p$. From the fact that $\{0, 1\}^p = \{(1_U, 0_{V \setminus U}) \mid U \subseteq V\}$, it follows that one can write the probability distribution of X_V as a vector $\pi = (\pi_U)_{U \subseteq V}$ with entries $\pi_U = P(X_U = 1_U, X_{V \setminus U} = 0_{V \setminus U})$.

The multivariate Bernoulli distribution belongs to the natural exponential family with mean parameter $\mu = (\mu_U)_{U \subseteq V}$ where

$$\mu_\emptyset = 1 \quad \text{and} \quad \mu_U = \text{pr}(X_U = 1_U) \quad \text{for every } U \subseteq V \text{ with } U \neq \emptyset. \quad (1)$$

The mean parameter μ was called the *Möbius parameter* by Drton and Richardson (2008) because $\mu_U = \sum_{E \subseteq V \setminus U} \pi_{U \cup E}$ for every $U \subseteq V$ so that the inverse map $\mu \mapsto \pi$ can be computed by *Möbius inversion* as $\pi_U = \sum_{E \subseteq V \setminus U} (-1)^{|E|} \mu_{U \cup E}$ for every $U \subseteq V$; see Lauritzen (1996, Appendix A). Let Z and M be two $(2^p \times 2^p)$ matrices with entries indexed by the subsets of $V \times V$ and given by $Z_{U,H} = 1(U \subseteq H)$ and $M_{U,H} = (-1)^{|H \setminus U|} 1(U \subseteq H)$, respectively, where $1(\cdot)$ denotes the indicator function. Then, one can write the linear relationship between π and μ in matrix form as $\mu = Z\pi$ and $\pi = M\mu$ and Möbius inversion follows by noticing that $M = Z^{-1}$.

Roverato et al. (2013) introduced the Log-Mean Linear (LML) parameter $\gamma = (\gamma_U)_{U \subseteq V}$ whose entries are computed as a log-linear expansion of the Möbius parameters

$$\gamma_U = \sum_{E \subseteq U} (-1)^{|U \setminus E|} \log \mu_E \quad \text{for every } U \subseteq V \quad (2)$$

so that, in matrix form, $\gamma = M^\top \log \mu$ where M^\top denotes the transpose of M . Note that $\gamma_\emptyset = 0$ and, furthermore, that π can be analytically computed by applying Möbius inversion

twice to obtain $\pi = M \exp Z^\top \gamma$. The LML parameter can be regarded as a linearization of the Möbius parameter because certain multiplicative constraints in μ correspond to LML interactions equal to zero as follows.

Theorem 3.1 (Theorem 1 in Roverato et al. (2013)). *Let X_V be a vector of binary variables with Möbius and LML parameters μ and γ , respectively. Then, for a pair of disjoint, nonempty, proper subsets A and B of V , the following conditions are equivalent:*

- (i) $X_A \perp\!\!\!\perp X_B$;
- (ii) $\mu_{A' \cup B'} = \mu_{A'} \times \mu_{B'}$ for every $A' \subseteq A$ and $B' \subseteq B$;
- (iii) $\gamma_{A' \cup B'} = 0$ for every $A' \subseteq A$ and $B' \subseteq B$ such that $A' \neq \emptyset$ and $B' \neq \emptyset$.

The equivalence (i) \Leftrightarrow (ii) of Theorem 3.1 follows immediately from Theorem 1 of Drton and Richardson (2008) where it is shown that, in the binary case, bi-directed graph models can be parameterized by imposing multiplicative constraints on the Möbius parameters. The latter result can be expressed in terms of the LML parameters as follows.

Theorem 3.2 (Theorem 2 in Roverato et al. (2013)). *Let $\gamma = (\gamma_U)_{U \subseteq V}$ be the LML parameter of the binary random vector X_V and let $\mathcal{G} = (V, E)$ be a bi-directed graph. The distribution of X_V satisfies the connected set Markov property with respect to \mathcal{G} if and only if for every set $U \subseteq V$ that is disconnected in \mathcal{G} it holds that $\gamma_U = 0$.*

Example 3.1 (*bi-directed 4-chain cont.*). The disconnected sets of the bi-directed graph of Figure 1 are $\{1, 3\}$, $\{1, 4\}$, $\{2, 4\}$, $\{1, 2, 4\}$ and $\{1, 3, 4\}$ so that X_V satisfies the connected set Markov property with respect to such graph if and only if $\gamma_{\{1,3\}} = \gamma_{\{1,4\}} = \gamma_{\{2,4\}} = \gamma_{\{1,2,4\}} = \gamma_{\{1,3,4\}} = 0$.

Theorem 3.1 and 3.2 are proved in Roverato et al. (2013) under the assumption that π is strictly positive, and this implies that γ_U is well-defined for every $U \subseteq V$. Here, it is worth remarking that, for every $U \subseteq V$, γ_U in (2) is well-defined if and only if $\mu_U > 0$ because $\mu_E \geq \mu_U$ for every $E \subseteq U$.

3.2 The general discrete case

In this section, we consider the general case where the variables in Y_V take on an arbitrary number of levels that we label as $I_v = \{0_v, 1_v, \dots, d_v\}$, for every $v \in V$. Hence, the state space of Y_V is the product set $\mathcal{I}_V = \times_{v \in V} I_v$ that, with a slight abuse of terminology, we call a *cross-classified table*. Accordingly, the elements $i \equiv i_V \in \mathcal{I}_V$ are called the *cells* of the table. The probability distribution of Y_V is characterized by the probability table $\varpi = (\varpi_i)_{i \in \mathcal{I}_V}$, which we assume to be strictly positive. We remark that the symbol ϖ is used to denote the probability table of an arbitrary discrete random vector whereas π is used only in the binary case.

In the general discrete case, two different parameterizations of bi-directed graph models are available. Lupporelli et al. (2009) showed that there exists a connection between bi-directed graph models and the *multivariate logistic parameterization* of Glonek and McCullagh (1995). Specifically, every multivariate logistic interaction is a log-linear parameter computed in the relevant marginal distribution and a bi-directed graph model can be specified by setting to zero all the interactions corresponding to the disconnected sets of the graph; see also Rudas et al. (2010) and Evans and Richardson (2012). Parsimony can be achieved by setting further interactions to zero, but such additional constraints are typically difficult to interpret.

Drton (2009) generalized the Möbius parameters to include non-binary variables and regression graph models. From the state space of Y_v , Drton (2009) introduced the *restricted state space* defined as $J_v = I_v \setminus \{0_v\} = \{1_v, \dots, d_v\}$ so that the restricted state space of Y_V is given by $\mathcal{J}_V = \times_{v \in V} J_v$. Hereafter, we refer to “ 0_v ” as to the *baseline* level of Y_v and remark that the choice of the level to be set as baseline is arbitrary. For every $U \subseteq V$, with $U \neq \emptyset$, we denote by \mathcal{I}_U and \mathcal{J}_U the state space and the restricted state space, respectively, of Y_U . Furthermore, for every $j \in \mathcal{J}_V$ we denote by j_U the subset of levels of j corresponding to the entries of Y_U , so that we can write $j_U \subseteq j$, and it holds that $\mathcal{J}_U = \{j_U \mid j \in \mathcal{J}\}$. Finally, when $U = \emptyset$ we use the convention that $j_U = \mathcal{J}_U = \emptyset$.

The *saturated Möbius parameter* of Y_V was defined by Drton (2009) as the collection of marginal probabilities

$$\text{pr}(Y_U = j_U) \quad \text{for every } j \in \mathcal{J}_V \text{ and } U \subseteq V \text{ with } U \neq \emptyset. \quad (3)$$

Drton (2009) showed the saturated Möbius parameters characterize the distribution of Y_V and that every bi-directed graph model is defined by an appropriate choice of multiplicative constraints on the saturated Möbius parameters. However, it is not clear how parsimony can be achieved with this parameterization. We also recall that the saturated Möbius parameters are closely related to the dependence ratios of Ekholm et al. (2000).

4 The LML parameterization for the general discrete case

In this section we extend the LML parameterization of Roverato et al. (2013) to discrete random variables with arbitrary number of levels and show that, also in this case, every bi-directed graph model can be defined by setting certain LML interactions to zero.

For every $v \in V$ and $i_v \in I_v$ we introduce the Bernoulli random variable $X_v^{i_v}$ defined as

$$X_v^{i_v} = \begin{cases} 1 & \text{if } Y_v = i_v \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

In this way, every cell $i \in \mathcal{I}_V$ is associated with the random vector $X_V^i = (X_v^{i_v})_{v \in V}$ which follows a multivariate Bernoulli distribution, and we denote by $\pi^i = (\pi_U^i)_{U \subseteq V}$ the corresponding probability parameter. Accordingly, the mean and LML parameter of X_V^i can be computed as $\mu^i = Z\pi^i$ and $\gamma^i = M^\top \log \mu^i$, respectively.

It is straightforward to see, from (1) and (4), that

$$\mu_U^i = \text{pr}(X_U^i = 1_U) = \text{pr}(Y_U = i_U) \quad \text{for every } i \in \mathcal{I}_V \text{ and } U \subseteq V \text{ with } U \neq \emptyset, \quad (5)$$

and it follows that the Möbius parameter in (3) can be written as a collection of vectors of Möbius parameters for binary vectors; formally μ^j for $j \in \mathcal{J}_V$. In this way one can see that the collection of LML parameters γ^j , for $j \in \mathcal{J}_V$, parameterizes the distribution of Y_V , because there exists a bijective map between γ^j and μ^j for every $j \in \mathcal{J}_V$. For every $i, i' \in \mathcal{I}_V$ and $U \subseteq V$ such that $i_U = i'_U$ it holds both that $\mu_U^i = \mu_U^{i'}$ and that $\gamma_U^i = \gamma_U^{i'}$ and, to remove redundancies, we write $\mu^{i_U} = \mu_U^i$ and $\gamma^{i_U} = \gamma_U^i$ so that the Möbius and the LML parameters of Y_V can be formally defined as

$$\mu = (\mu^{j_U})_{j \in \mathcal{J}_V, U \subseteq V} \quad \text{and} \quad \gamma = (\gamma^{j_U})_{j \in \mathcal{J}_V, U \subseteq V},$$

respectively.

The fact that both the Möbius and the LML parameter for the general case are defined as the collection of Möbius and LML parameters, respectively, for the collection of binary vectors X_V^j with $j \in \mathcal{J}_V$ represents a key feature that confers useful properties to our approach. For instance, the generalization of relevant properties for these parameterizations follows immediately from the iterative application, for every $j \in \mathcal{J}_V$, of the corresponding properties of binary vectors. This is the case of the result of Roverato et al. (2013), given in Theorem 3.1.

Theorem 4.1. *Let μ and γ the Möbius and LML parameters of Y_V , respectively. Then for a pair of disjoint, nonempty, proper subsets A and B of V , the following conditions are equivalent:*

- (i) $Y_A \perp\!\!\!\perp Y_B$;
- (ii) $\mu^{j_{A' \cup B'}} = \mu^{j_{A'}} \times \mu^{j_{B'}}$ for every $j \in \mathcal{J}_V$, $A' \subseteq A$ and $B' \subseteq B$;
- (iii) $\gamma^{j_{A' \cup B'}} = 0$ for every $j \in \mathcal{J}_V$, $A' \subseteq A$ and $B' \subseteq B$ such that $A' \neq \emptyset$ and $B' \neq \emptyset$.

Proof. See the Appendix A. □

Theorem 4.1 can be applied to show that bi-directed graph models can be parameterized by setting to zero the LML interactions corresponding to the disconnected sets of \mathcal{G} , generalizing in this way the result of Roverato et al. (2013) given in Theorem 3.2.

Corollary 4.2. *Let $\gamma = (\gamma^{j_U})_{j \in \mathcal{J}_V, U \subseteq V}$ be the LML parameter of Y_V and let $\mathcal{G} = (V, E)$ be a bi-directed graph. The distribution of Y_V satisfies the connected set Markov property with respect to \mathcal{G} if and only if for every set $U \subseteq V$ that is disconnected in \mathcal{G} it holds that $\gamma^{j_U} = 0$ for every $j_U \in \mathcal{J}_U$.*

Proof. See the Appendix A. □

Example 4.1 (*bi-directed 4-chain cont.*). As in Example 3.1, we consider the graph \mathcal{G} in Figure 1. Then, by Corollary 4.2, the distribution of Y_V satisfies the connected set Markov property with respect to \mathcal{G} if and only if it holds that $\gamma^{j_U} = 0$ for every $U \in \{\{1, 3\}, \{1, 4\}, \{2, 4\}, \{1, 2, 4\}, \{1, 3, 4\}\}$ and $j_U \in \mathcal{J}_U$, i.e., if and only if $\gamma^{j_{\{1,3\}}} = \gamma^{j_{\{1,4\}}} = \gamma^{j_{\{2,4\}}} = \gamma^{j_{\{1,2,4\}}} = \gamma^{j_{\{1,3,4\}}} = 0$ for every $j \in \mathcal{J}_V$. Clearly, if Y_V is binary then $|\mathcal{J}_V| = 1$ and the zero LML interactions are the same as in Example 3.1.

The *discrete bi-directed graph model* for Y_V with graph $\mathcal{G} = (V, E)$ is defined as the set of positive probability distributions for Y_V that obey the connected set Markov property with respect to \mathcal{G} . Drton (2009, Corollary 10) showed that discrete bi-directed graph models are curved exponential families. The simple nature of the mapping $\varpi \mapsto \gamma$ allows one to see that γ is a smooth parameterization of the saturated model and therefore that any model defined by imposing linear constraints on the LML parameter of the saturated model is a curved exponential family. The family of submodels defined by linear constraints on γ includes discrete bi-directed graph models by Corollary 4.2, and in the next section we will see how additional zero constraints on the LML parameters can be specified so as to obtain interpretable bi-directed graph submodels.

Maximum likelihood estimation for LML models under a multinomial or Poisson sampling scheme is a constrained optimization problem, that can be carried out by using gradient-based ascent methods; we refer to Lang (1996) and Bergsma et al. (2009) for details. We remark that the likelihood function can be expressed in terms both of Möbius parameters and of LML parameters but not analytically in terms of multivariate logistic parameters because an analytic form of the inverse map to compute ϖ from the marginal logistic parameters is not available; see also Roverato et al. (2013).

5 Dichotomization invariance and B -expanded graphs

The LML parameterization is based on the collection of binary vectors X_V^j for $j \in \mathcal{J}$ and it is useful to take a closer look at how these parameters are computed. For every $i \in \mathcal{I}_V$ the computation of μ^i is based on the probability table π^i of X_V^i and one should observe that π^i is obtained by collapsing the levels of Y_V in such a way that the probability in the cell $i \in \mathcal{I}_V$ is not affected by this operation; formally, $\varpi_i = \pi_V^i$. For this reason, we say that π^i is constructed by collapsing the levels of Y_V “around” the cell i , and this trivially implies that the value of μ^i , as well as that of γ^i , is unaffected by collapsing operations that do not involve the level i_v for every $v \in V$. As a consequence of this invariance property of the LML parameterization, certain zero entries in the LML parameter γ of Y_V allow one to identify marginal independencies concerning dichotomized versions of the variables, as well as to identify levels of the variables that can be collapsed without affecting the structure of the associated bi-directed graph. We formally approach this issue by introducing the concept of binary expansion of a discrete variable, that is based on the dichotomization (4).

Definition 5.1. For $v \in V$, the *binary expansion of Y_v with respect to J_v* is the $|J_v|$ -dimensional random vector of binary variables $X_{J_v} = (X_v^{j_v})_{j_v \in J_v}$.

The binary expansion X_{J_v} provides an alternative representation of Y_v : every entry of X_{J_v} corresponds to a different dichotomization of Y_v so that, for every $j_v \in J_v$, the variable $X_v^{j_v}$ takes on the value 1 if and only if $Y_v = j_v$ and value 0 otherwise. Moreover, $X_{J_v} = 0_{J_v}$ if and only if $Y_v = 0_v$. Clearly, there exist $|I_v|$ different binary expansions of Y_v , depending on the choice of the baseline level, and they are all equivalent, in the sense that there is a one-to-one relationship between Y_v and any of its binary expansions. On the other hand, the specification of one binary expansion, out of the $|I_v|$ existing alternatives, amounts to fixing a particular perspective from which the variable structure is explored. A suitable choice of the baseline level may correspond to a binary expansion of special interest and, ultimately, make it possible to disclose relevant additional features concerning the association structure of the variables.

Example 5.1 (*Genetic association analysis cont.*). Let Y_v be the variable representing a given SNP under the codominant genotype model so that Y_v takes values in the set $\{WM, WW, MM\}$. There exist three different binary expansions of Y_v . However, if one sets the genotype WM as baseline level, then every entry of the resulting binary expansion X_{J_v} has a clear interpretation. Indeed, one of the two entries is associated with WW and therefore it corresponds to the “dominant” dichotomization $WM + MM$ vs. WW whereas the other entry is associated with MM and corresponds to the “recessive” dichotomization $WM + WW$ vs. MM . In order to make our notation more intuitive, hereafter for a SNP Y_v we label WW as D_v (for Dominant), and MM as R_v (for Recessive) so that $J_v = \{D_v, R_v\}$ and $X_{J_v} = (X_v^{D_v}, X_v^{R_v})^\top$.

We now turn to the variable relative to the phenotype or trait of interest. It is common for a genetic association study to be based on a case-control design where the phenotype Y_v is a discrete variable with levels corresponding to the controls and to the different states of a given disease for the cases. Hence, by setting the controls as baseline level, every entry of the resulting binary expansion X_{J_v} corresponds to one of the different states of the disease.

The concept of binary expansion of a variable can then be extended to that of binary expansion of a random vector Y_B , $B \subseteq V$, with respect to $J_B = \cup_{v \in B} J_v$ that is given by $X_{J_B} = (X_v^{j_v})_{j_v \in J_v, v \in B}$. In the rest of the paper, we assume, without loss of generality, that J_B is fixed and shortly write that X_{J_B} is the binary expansion of Y_B . Furthermore, if $P = V \setminus B$ then we write $Y_V^B = (Y_P, X_{J_B})$ and say that Y_V^B is the *B-expansion* of Y_V ; note that $Y_V^V = X_{J_V}$ whereas, if $B = \emptyset$ then $J_B = \emptyset$ and $Y_V^B = Y_V$.

The main result of this section is a generalization of Theorem 4.2 where we show that, for every $B \subseteq V$, bi-directed graph models for Y_V^B can be parameterized by setting certain entries of γ to zero. We emphasize that here γ is the LML parameter of Y_V and therefore the subset B plays no role in its computation.

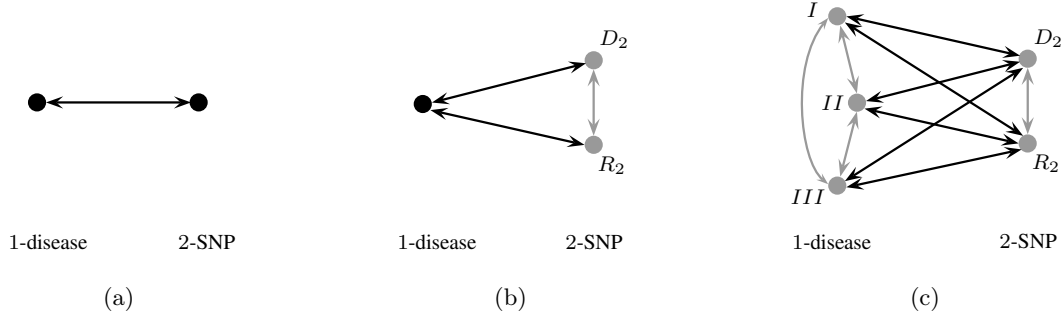


Figure 2: Example of complete expanded graphs for variables Y_1 and Y_2 with $J_1 = \{I, II, III\}$ and $J_2 = \{D_2, R_2\}$.

The marginal independence of $Y_V^B = (Y_P, X_{J_B})$ is encoded by a graph on $P \cup J_B$ vertices. By construction, for every $v \in V$, no marginal independence is present between the entries of X_{J_v} because $X_v^{j_v} = 1$ implies that $X_{J_v \setminus \{j_v\}} = 0_{J_v \setminus \{j_v\}}$, for every $j_v \in J_v$. Hence, the bi-directed graph of Y^B is a B -expanded graph, formally defined below.

Definition 5.2. let Y_V be a discrete random variable and $V = P \dot{\cup} B$ a partition of V . We say that \mathcal{G}^B is a B -expanded graph for Y_V if it is a bi-directed graph with vertex set equal to $P \cup J_B$, and such that the subgraph induced by J_v is complete for every $v \in B$.

Example 5.2 (*Genetic association analysis cont.*). Consider a case-control genetic association study where Y_1 is a variable whose baseline level corresponds to the controls and $J_1 = \{I, II, III\}$ encodes three different stages of a disease measured on the cases. Furthermore, let Y_2 be a SNP and $J_2 = \{D_2, R_2\}$. The graphs in Figure 2 represent: (a) the complete B -expanded graph with $B = \emptyset$, that is, the complete bi-directed graph with vertex set $\{1, 2\}$, (b) the complete B -expanded graph with $B = \{2\}$ and (c) the complete B -expanded graph with $B = \{1, 2\}$. Note that, to improve readability, the gray color is used to represent the complete subgraphs of the expanded variables.

Before stating the connected set Markov property on B -expanded graphs it is convenient to introduce the notion of primary subsets.

Definition 5.3. Let $J_B = \cup_{v \in B} J_v$ where $B \subseteq V$ and $P = V \setminus B$. We say that K is a *primary* subset of J_B if $K \subseteq J_B$ and it contains at most one level for every variable in Y_B ; formally, $|K \cap J_v| \leq 1$ for every $v \in B$. Furthermore, we say that L is a primary subset of $P \cup J_B$ if $L \subseteq P \cup J_B$ and $K = L \cap J_B$ is a primary subset of J_B .

Note that the empty set is always primary and, furthermore, if $B = \emptyset$ then $P = V$ and, in this case, every subset of P is primary.

Lemma 5.1. Every primary subset L of $P \cup J_B$ can be partitioned as $L = Q \dot{\cup} K$ where $Q = P \cap L$ and $K = J_B \cap L$. Moreover, there exists a unique subset $D \subseteq B$ such that

$K \in \mathcal{J}_D$ and, for this reason, we can write $K = j_D$ and $L = Q \cup j_D$. Conversely, for every $Q \subseteq P$, $D \subseteq B$ and $j_D \in \mathcal{J}_D$ it holds that $Q \cup j_D$ is a primary subset of $P \cup J_B$.

Proof. This is a straightforward consequence of the fact that if L is primary and $K = L \cap J_B$ then $|K \cap J_v| \leq 1$ for every $v \in B$. \square

We can now state the main result of this section.

Theorem 5.2. *Let $\gamma = (\gamma^{jv})_{j \in \mathcal{J}_V, v \subseteq V}$ be the LML parameter of Y_V and let $\mathcal{G}^B = (P \cup J_B, E^B)$ be a B -expanded graph for Y_V . The distribution of $Y^B = (Y_P, X_{J_B})$ satisfies the connected set Markov property with respect to \mathcal{G}^B if and only if for every set $L \subseteq P \cup J_B$ such that*

- (i) L is disconnected in \mathcal{G}^B ,
- (ii) L is a primary subset of $P \cup J_B$,

it holds that $\gamma^{j_{Q \cup D}} = 0$ for every $j_Q \in \mathcal{J}_Q$, where $L = Q \cup j_D$ as in Lemma 5.1 and $j_{Q \cup D} = j_Q \cup j_D$.

Proof. See the Appendix A. \square

The following example clarifies the connection between LML parameters and edges of expanded graphs in the simple case where only two variables are considered. Interestingly, when both variables are expanded, every edge of the expanded graph can be associated with exactly one LML parameter.

Example 5.3 (*Genetic association analysis cont.*). Let Y_1 and Y_2 be the two variable depicted in the graphs of Figure 2. The LML parameter of $Y_{\{1,2\}}$ is made up of: $\gamma^\emptyset = 0$; the main effects of Y_1 that are $\gamma^{\{I\}}$, $\gamma^{\{II\}}$ and $\gamma^{\{III\}}$; the main effects of Y_2 given by $\gamma^{\{D_2\}}$ and $\gamma^{\{R_2\}}$ and, finally, the two-way interactions $\gamma^{\{I,D_2\}}$, $\gamma^{\{II,D_2\}}$, $\gamma^{\{III,D_2\}}$, $\gamma^{\{I,R_2\}}$, $\gamma^{\{II,R_2\}}$ and $\gamma^{\{III,R_2\}}$. It follows from Corollary 4.2 that the edge $1 \leftrightarrow 2$ in the graph (a) of Figure 2 can be removed if and only if all of the six two-way LML interactions are equal to zero, so that $Y_1 \perp\!\!\!\perp Y_2$. Consider now the graph (b) in the same figure. Here, there are only two primary subsets involving more than one vertex, $\{1, D_2\}$ and $\{1, R_2\}$. The edge $1 \leftrightarrow D_2$ encodes the marginal association between Y_1 and the dominant version $X_2^{D_2}$ of the SNP Y_2 and, by Theorem 5.2, it can be removed if and only if $\gamma^{\{I,D_2\}} = \gamma^{\{II,D_2\}} = \gamma^{\{III,D_2\}} = 0$. Similarly, the edge $1 \leftrightarrow R_2$ can be removed if and only if $\gamma^{\{I,R_2\}} = \gamma^{\{II,R_2\}} = \gamma^{\{III,R_2\}} = 0$. It is therefore clear that Y_1 is independent of the codominant representation Y_2 of the SNP if and only if it is independent both of the dominant representation $X_2^{D_2}$ and the recessive representation $X_2^{R_2}$. Of main interest is the case where only one edge is missing in the graph (b) because the B -expanded provides additional insight into the independence structure of the two variables with respect to the traditional graph for $Y_{\{1,2\}}$. We now turn to the B -expanded graph (c) in Figure 2. In this case, $P = \emptyset$, $J_B = \{I, II, III, D_2, R_2\}$ and the

primary subsets of J_B which involve more than one vertex are $\{I, D_2\}$, $\{II, D_2\}$, $\{III, D_2\}$, $\{I, R_2\}$, $\{II, R_2\}$ and $\{III, R_2\}$. Hence, it follows from Theorem 5.2 that every edge of the graph can be removed if and only if the corresponding two way interaction is equal to zero. For instance, the edge $II \leftrightarrow R_2$ can be removed if and only if $\gamma^{\{II, R_2\}} = 0$.

We now illustrate the potentiality of B -expanded graphs to provide interpretable parsimonious bi-directed graph submodels.

Example 5.4 (*Genetic association analysis cont.*). Let $V = \{1, 2, 3\}$ where Y_1 and Y_2 are the two variable in the graphs in Example 5.4 and Y_3 is an additional SNP. In this case, apart from $\gamma^\emptyset = 0$, the LML parameter is made up of 35 entries, concretely: 7 main effects, 16 two-way interactions and 12 three-way interactions. Assume that the probability distribution of Y_V is such that the following 18 LML parameters are equal to zero: $\gamma^{\{j_1, R_2\}}$, $\gamma^{\{j_1, R_3\}}$, $\gamma^{\{D_2, D_3\}}$, $\gamma^{\{R_2, D_3\}}$, $\gamma^{\{D_2, R_3\}}$, $\gamma^{\{j_1, R_2, R_3\}}$, $\gamma^{\{j_1, R_2, D_3\}}$ and $\gamma^{\{j_1, D_2, R_3\}}$ for every $j_1 \in J_1$. It is easy to see from Theorem 3.1 that in this case there are no pairwise marginal independencies so that the distribution of Y_V is associated with the complete graph (a) in Figure 3. However, the distribution of Y_V belongs to a parsimonious model that can be completely defined in terms of marginal independence relationships. If we set $B = \{2, 3\}$, then the zero LML parameters above are associated with the subsets of $\{1\} \cup J_B$ with which are both primary and disconnected in the graph (b) of Figure 3. Hence, by Theorem 5.2, the distribution of Y_V^B satisfies the connected set Markov property with respect to the latter graph that implies, among others, the independence of the disease of the recessive versions of the SNPs; $Y_1 \perp\!\!\!\perp (X_2^{R_2}, X_3^{R_3})$. Clearly, it would be possible also to expand Y_1 , but this would make the graph unnecessarily more complex because the zero structure of γ does not allow us to remove any edge with an endpoint in the expansion of this variable. However, if additional interactions are equal to zero such as, for instance, $\gamma^{\{III, D_2\}}$, $\gamma^{\{II, D_3\}}$, $\gamma^{\{III, D_3\}}$, $\gamma^{\{III, D_2, D_3\}}$ and $\gamma^{\{II, D_2, D_3\}}$, then it makes sense to expand also Y_1 so as to obtain the B -expanded graph (c) in Figure 3. This graph encodes marginal relationships involving single levels of Y_1 , such as $X_1^{III} \perp\!\!\!\perp (X_2^{R_2}, X_2^{D_2}, X_3^{R_3}, X_3^{D_3})$ that is equivalent to $X_1^{III} \perp\!\!\!\perp Y_{\{2, 3\}}$. Note that, in the latter model, 23 out of the 35 LML parameters are constrained to zero but the independence structure of Y_V is still represented by the complete graph.

The rest of this section is devoted to some basic results that are required for the proof of Theorem 5.2. These are rather technical but, nevertheless, of interest because they provide a formal proof of the dichotomization invariance property of both the Möbius and the LML parameterizations. It is worth remarking that most of the existing results for marginal independence models, including Theorem 4.1 and Corollary 4.2, cannot be directly applied to Y_V^B because its distribution contains structural zeros. In fact, the following lemma shows that the Markov property for B -expanded graphs is characterized by the same property for the collection of subvectors (Y_P, X_{j_B}) , for $j_B \in \mathcal{J}_B$, whose probability distributions have no structural zeros.

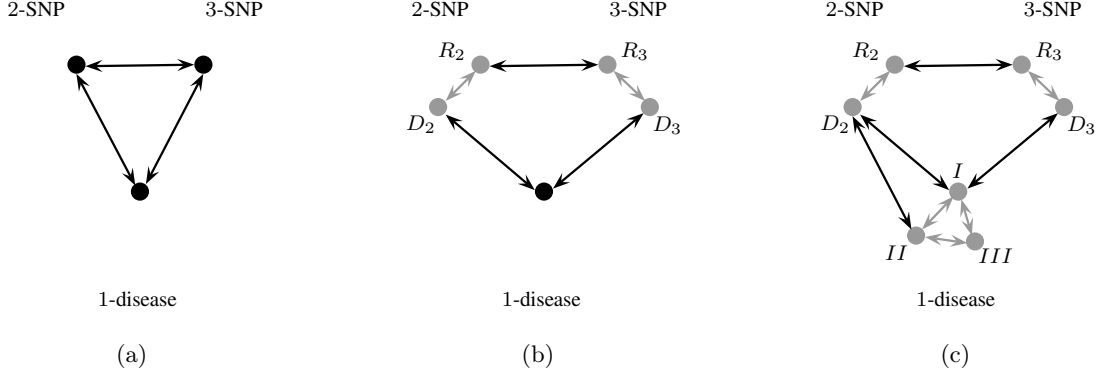


Figure 3: Example of expanded graphs for $Y_{\{1,2,3\}}$ with $J_1 = \{I, II, III\}$ and $J_i = \{D_i, R_i\}$ for $i = 2, 3$.

Lemma 5.3. *Let $\mathcal{G}^B = (P \cup J_B, E^B)$ be a B -expanded graph for Y_V . The distribution of $Y^B = (Y_P, X_{J_B})$ satisfies the connected set Markov property with respect to \mathcal{G}^B if and only if the distribution of (Y_P, X_{j_B}) satisfies the connected set Markov property with respect to $\mathcal{G}_{P \cup j_B}^B$ for every $j_B \in \mathcal{J}_B$.*

The probability table of Y_B is made up of $|\mathcal{I}_B|$ positive probabilities that can be written as $\text{pr}(Y_B = i_B)$ for every $i_B \in \mathcal{I}_B$ or, equivalently, as $\text{pr}(Y_D = j_D, Y_{B \setminus D} = 0_{B \setminus D})$ for every $j_D \in \mathcal{J}_D$ and $D \subseteq B$. We have named X_{J_B} the binary expansion of Y_B because its probability table can be hypothetically constructed by adding $2^{|J_B|} - |\mathcal{I}_B|$ structural-zero cells to the probability table of Y_B . More specifically, the probability distribution of X_{J_B} can be written as $\text{pr}(X_K = 1_K, X_{J_B \setminus K} = 0_{J_B \setminus K})$, for $K \subseteq J_B$ and, these probabilities are equal to zero if and only if K is non-primary.

Lemma 5.4. *Let X_{J_B} be the binary expansion of Y_B and let $K \subseteq J_B$. If K is a primary subset of J_B then we can write $K = j_D$ by Lemma 5.1 and it holds both that $\text{pr}(X_K = 1_K, X_{J_B \setminus K} = 0_{J_B \setminus K}) = \text{pr}(Y_D = j_D, Y_{B \setminus D} = 0_{B \setminus D})$ and, if $j_D \neq \emptyset$, that $\text{pr}(X_K = 1_K) = \text{pr}(Y_D = j_D)$. Furthermore, if K is non-primary then $\text{pr}(X_K = 1_K, X_{J_B \setminus K} = 0_{J_B \setminus K}) = \text{pr}(X_K = 1_K) = 0$.*

Proof. See the Appendix A. □

We turn now to $Y_V^B = (Y_P, X_{J_B})$. The Möbius and the LML parameters of Y_V are $\mu = (\mu^{j_V})_{j \in \mathcal{J}_V, U \subseteq V}$ and $\gamma = (\gamma^{j_V})_{j \in \mathcal{J}_V, U \subseteq V}$, respectively. Furthermore, we denote the Möbius and the LML parameters of Y_V^B by $\tilde{\mu}$ and $\tilde{\gamma}$, respectively, where the entries of $\tilde{\mu}$ are $\tilde{\mu}^{j_Q \cup 1_K}$ for every $j_Q \in \mathcal{J}_Q$ and $K \subseteq J_B$, and we use the shorthand $\tilde{\mu}^{j_Q \cup 1_K} = \tilde{\mu}_K^{j_Q}$. Similarly, we denote the entries of $\tilde{\gamma}$ by $\tilde{\gamma}_K^{j_Q}$ for $j_Q \in \mathcal{J}_Q$ and $K \subseteq J_B$. The following theorem states that both the Möbius and the LML parameters are dichotomization invariant in the sense that μ and γ are subvectors of $\tilde{\mu}$ and $\tilde{\gamma}$, respectively, whereas the remaining entries of $\tilde{\mu}$ and $\tilde{\gamma}$ are equal to zero and not well-defined, respectively.

Theorem 5.5. *Let $Y_V^B = (Y_P, X_{J_B})$ be the B -expansion of Y_V . Furthermore, let $\tilde{\mu}_K^{j_Q}$ and $\tilde{\mu}_K^{j_Q}$, for $j_Q \in \mathcal{J}_Q$ and $K \subseteq J_B$, the Möbius and the LML parameter of Y^B , respectively, whereas μ and γ are the Möbius and the LML parameter of Y_V , respectively. If L is a primary subset of $P \cup J_B$, so that $L = Q \cup j_D$ as in Lemma 5.1, then $\mu^{j_{Q \cup D}} = \tilde{\mu}_{j_D}^{j_Q}$ and $\gamma^{j_{Q \cup D}} = \tilde{\gamma}_{j_D}^{j_Q}$, for every $j_Q \in \mathcal{J}_Q$. Conversely, if L is non-primary and we set $K = L \cap J_B$ and $Q = L \cap P$, then $\tilde{\mu}_K^{j_Q} = 0$ and, consequently, $\tilde{\gamma}_K^{j_Q}$ is not well-defined, for every $j_Q \in \mathcal{J}_Q$.*

Proof. See the Appendix A. □

6 Discussion

Roverato et al. (2013) introduced the LML parameterization for binary data and explained its advantages with respect to both the multivariate logistic and the Möbius parameterization. The extension of this approach to variables with an arbitrary number of levels, has disclosed the additional invariance property that makes it possible to use the LML parameter of Y_V to characterize the connected set Markov property for any B -expansion Y_V^B of Y_V . As a consequence, using the LML parameterization amounts to implicitly working with the binary expansions of variables or, from a different perspective, the LML parameterization allows one to deal with expanded variables implicitly, so as to avoid all the difficulties associated with the expansion operation. For instance, one has not to worry either for the presence of structural zeros or for the inefficiencies deriving from the artificial increase in dimensionality. In structural learning, the set B does not need to be defined a priori, but it can be specified after a LML model has been selected from data. For instance, B can be chosen so as to optimize the trade-off between readability of the graph and the need to explicit the learnt independencies involving every single expanded variable.

Open questions include the specification of the baseline level in situations where there is no ‘natural’ baseline level, as well as the specification of the baseline level for ordinal variables. The general issue of developing model search strategies for the identification of a parsimonious bi-directed graph submodel is still open. However, it may be appealing to restrict the search space to the models characterized by the family of V -expanded graphs, because it is made up of interpretable models, and it is smaller than the family of all the models obtained by constraining a subset of the LML parameters to zero.

We close this discussion by remarking that, by Theorem 5.5, also the Möbius parameterization satisfies the same dichotomization invariance property as the LML parameterization. However, the LML parameterization has the advantage that marginal independence relationships correspond to, linear, zero constraints in the space of the parameters.

Acknowledgments

We gratefully acknowledge useful discussions with Robert Castelo, Monia Lupparelli and Luca la Rocca.

A Proofs

Proof of Theorem 4.1

(i) \Rightarrow (ii). Firstly, we note that the factorizations in (ii) are trivially true when at least one between A' and B' is equal to the empty set because $\mu^{j\emptyset} = 1$. The independence $Y_A \perp\!\!\!\perp Y_B$ implies that $Y_{A'} \perp\!\!\!\perp Y_{B'}$ for every $A' \subseteq A$ and $B' \subseteq B$ with $A', B' \neq \emptyset$. In turn, $Y_{A'} \perp\!\!\!\perp Y_{B'}$ implies that, for every $j \in \mathcal{J}_V$, $\text{pr}(Y_{A'} = j_{A'}, Y_{B'} = j_{B'}) = \text{pr}(Y_{A'} = j_{A'}) \times \text{pr}(Y_{B'} = j_{B'})$ and since $Y_{A'} = j_{A'}$ if and only if $X_{A'}^j = 1_{A'}$, and similarly for $Y_{B'}$, this implies that $\text{pr}(X_{A'}^j = 1_{A'}, X_{B'}^j = 1_{B'}) = \text{pr}(X_{A'}^j = 1_{A'}) \times \text{pr}(X_{B'}^j = 1_{B'})$ and the result follows by (5) because the latter factorization can be written as $\mu^{j_{A' \cup B'}} = \mu^{j_{A'}} \times \mu^{j_{B'}}$.

(i) \Leftarrow (ii). By Theorem 3.1, (ii) implies that $X_A^j \perp\!\!\!\perp X_B^j$ for every $j \in \mathcal{J}_V$, and therefore that $\text{pr}(X_A^j = 1_A, X_B^j = 1_B) = \text{pr}(X_A^j = 1_A) \times \text{pr}(X_B^j = 1_B)$, for every $j \in \mathcal{J}_V$. The latter factorization is equivalent to

$$\text{pr}(Y_A = j_A, Y_B = j_B) = \text{pr}(Y_A = j_A) \times \text{pr}(Y_B = j_B) \quad \text{for every } j \in \mathcal{J}_V \quad (6)$$

and we have to show that the factorization in (6) holds for every $i \in \mathcal{I}_V$. This follows immediately from Theorem 8 of Drton (2009) but we formally prove it for completeness. Every $i \in \mathcal{I}_V$ with $i \notin \mathcal{J}_V$ contains at least one level labeled as baseline, that is as “0”, and the proof is by induction on the number baseline levels in i . Let k denote the number of baseline levels in $i \in \mathcal{I}_V$; the factorization in (6) holds for $k = 0$ and we show that if it is true for $k = r - 1$, with $r > 0$, then it is also true for $k = r$. Assume that $k = r$ and that, without loss of generality, $Y_v = 0_v$ with $v \in A$. Hence, if $A' = A \setminus \{v\}$,

$$\begin{aligned} \text{pr}(Y_A = i_A, Y_B = i_B) &= \text{pr}(Y_v = 0_v, Y_{A'} = i_{A'}, Y_B = i_B) \\ &= \text{pr}(Y_{A'} = i_{A'}, Y_B = i_B) - \sum_{i_v=1_v}^{d_v} \text{pr}(Y_v = i_v, Y_{A'} = i_{A'}, Y_B = i_B) \end{aligned}$$

and, since the number of “0”’s in $i_{A' \cup B}$ is equal to $r - 1$, it follows from the induction assumption that

$$\begin{aligned} \text{pr}(Y_A = i_A, Y_B = i_B) &= \text{pr}(Y_{A'} = i_{A'}) \times \text{pr}(Y_B = i_B) - \sum_{i_v=1_v}^{d_v} \text{pr}(Y_v = i_v, Y_{A'} = i_{A'}) \times \text{pr}(Y_B = i_B) \\ &= \left\{ \text{pr}(Y_{A'} = i_{A'}) - \sum_{i_v=1_v}^{d_v} \text{pr}(Y_v = i_v, Y_{A'} = i_{A'}) \right\} \text{pr}(Y_B = i_B) \\ &= \text{pr}(Y_A = i_A) \times \text{pr}(Y_B = i_B) \end{aligned}$$

as required.

(ii) \Leftrightarrow (iii). This follows immediately from the equivalence, for every $j \in \mathcal{J}_V$, of (ii) and (iii) in Theorem 3.1.

Proof of Corollary 4.2

As a consequence of Theorem 4 of Richardson (2003), we have to show that $Y_{C_1} \perp\!\!\!\perp \cdots \perp\!\!\!\perp Y_{C_r}$ for every disconnected set U of \mathcal{G} if and only if for every disconnected set U of \mathcal{G} it holds that $\gamma^{j_U} = 0$ for every $j_U \in \mathcal{J}_U$. Recall that C_1, \dots, C_r are the connected components of U and that $r \geq 2$.

If $Y_{C_1} \perp\!\!\!\perp \cdots \perp\!\!\!\perp Y_{C_r}$ then we can set $A = C_1$ and $B = C_2 \dot{\cup} \cdots \dot{\cup} C_r$ so that $U = A \dot{\cup} B$ and $Y_A \perp\!\!\!\perp Y_B$. The result follows by noticing that every $j_U \in \mathcal{J}_U$ can be written as $j_{A \cup B}$ so that $\gamma^{j_U} = \gamma^{j_{A \cup B}} = 0$ by Theorem 4.1.

We now show the reverse implication, that is, that if $\gamma^{j_U} = 0$ for every $j_U \in \mathcal{J}_U$ such that U is disconnected in \mathcal{G} , then for every disconnected set U of \mathcal{G} it holds that $Y_{C_1} \perp\!\!\!\perp \cdots \perp\!\!\!\perp Y_{C_r}$. Let A and B be defined as above. Then, for every $A' \subseteq A$ and $B' \subseteq B$, with $A', B' \neq \emptyset$ the set $A' \cup B'$ is disconnected in \mathcal{G} so that, by assumption, $\gamma^{j_{A' \cup B'}} = 0$ for every $j_{A' \cup B'} \in \mathcal{J}_{A' \cup B'}$. This is equivalent to saying that $\gamma^{j_{A' \cup B'}} = 0$ for every $j \in \mathcal{J}_V$, $A' \subseteq A$ and $B' \subseteq B$ such that $A' \neq \emptyset$ and $B' \neq \emptyset$. The latter, by Theorem 4.1, implies that $Y_{A'} \perp\!\!\!\perp Y_{B'}$ or, equivalently, that $Y_{C_1} \perp\!\!\!\perp Y_{C_2 \cup \cdots \cup C_r}$. The same procedure can then be applied, for every $i = 1, \dots, r-1$, with $A = C_i$ and $B = C_{i+1} \cup \cdots \cup C_r$ to show that $Y_{C_i} \perp\!\!\!\perp Y_{C_{i+1} \cup \cdots \cup C_r}$ for every $i = 1, \dots, r-1$ which, in turn, implies that $Y_{C_1} \perp\!\!\!\perp \cdots \perp\!\!\!\perp Y_{C_r}$, as required.

Proof of Lemma 5.3

In this proof we repeatedly use the fact that if the distribution of a set of variables satisfies the connected set Markov property with respect to a given bi-directed graph then the marginal distribution of any subset of the variables satisfies the same Markov property with respect to the corresponding induced subgraph. This result follows immediately from the definition of the connected set Markov property because all the independence relationships implied by the subgraph are also encoded by the larger bi-directed graph.

Assume that the distribution of (Y_P, X_{J_B}) satisfies the connected set Markov property with respect to \mathcal{G}^B . Hence, for every $j_B \in \mathcal{J}_B$, the random vector (Y_P, X_{j_B}) is a subvector of (Y_P, X_{J_B}) and, therefore, its distribution satisfies the connected set Markov property with respect to the relevant induced subgraph $\mathcal{G}_{P \cup j_B}^B$. Hence, it is sufficient to show the inverse implication.

We denote by $L \subseteq P \cup J_B$ an arbitrary disconnected set in \mathcal{G}^B and by C_1, \dots, C_r its connected components. Furthermore, we set $Q = L \cap P$, $K = L \cap J_B$ and, to make the notation simpler, we write $Z = Y^B = (Y_P, X_{J_B})$. Hence, we have to show that if the distribution of (Y_P, X_{j_B}) satisfies the connected set Markov property with respect to $\mathcal{G}_{P \cup j_B}^B$

for every $j_B \in \mathcal{J}_B$ then

$$\text{pr}(Z_L = z_L) = \prod_{l=1}^r \text{pr}(Z_{C_l} = z_{C_l}) \quad \text{for every } z_L \in \mathcal{I}_Q \times \{0, 1\}^{|K|}. \quad (7)$$

For $L \subseteq P \cup J_B$ we set $D^* = \{v \mid v \in B \text{ and } K \cap J_v \neq \emptyset\}$. Note that, if L is primary we can write $L = Q \cup j_D$ by Lemma 5.1 and, in this case, $D = D^*$. Furthermore, it is easy to check that $|K| \geq |D^*|$ with $|K| = |D^*|$ if and only if L is primary. The proof is by induction on the value $(|K| - |D^*|)$.

We first assume $(|K| - |D^*|) = 0$, which includes the case where $K = \emptyset$. If $(|K| - |D^*|) = 0$ then $L = Q \cup j_D$ is primary and there exists a $j_B \in \mathcal{J}_B$ such that $L = Q \cup j_D \subseteq P \cup j_B$. Hence, the factorization (7) follows from the fact that, by assumption, (Y_P, X_{j_B}) satisfies the connected set Markov property with respect to $\mathcal{G}_{P \cup j_B}^B$.

We now show that if the result is true for $(|K| - |D^*|) = k$ with $k \geq 0$ then it is also true for $(|K| - |D^*|) = k + 1$. If $(|K| - |D^*|) = k + 1$ then $(|K| - |D^*|) > 0$ so that L is non-primary and, in turn, this implies that there exists a $v \in B$ such that $|J_v \cap K| \geq 2$. Hence, we set $J'_v = J_v \cap K$ and remark that the subgraph induced by J'_v is complete because \mathcal{G}^B is a B -expanded graph. This implies that J'_v is contained in exactly one connected component of L and we assume, without loss of generality, that $J'_v \subseteq C_1$. In this way, $X_{J'_v}$ is a subvector of Z_{C_1} which, in turn, is a subvector of Z_L . If $z_L \in \mathcal{I}_Q \times \{0, 1\}^{|K|}$ then there are three possible cases: (i) two or more variables in $X_{J'_v}$ take value 1, (ii) exactly one variable in $X_{J'_v}$ takes value 1 and (iii) all the variables in $X_{J'_v}$ take on the value 0.

In the case (i) we can find two distinct elements $j_v, j'_v \in J'_v$ such that $X_v^{j_v} = 1$ and $X_v^{j'_v} = 1$. However, $\text{pr}(X_v^{j_v} = 1, X_v^{j'_v} = 1) = \text{pr}(Y_v = j_v, Y_v = j'_v) = 0$ and this implies that both $\text{pr}(Z_L = z_L) = 0$ and $\text{pr}(Z_{C_1} = z_{C_1}) = 0$ so that the factorization (7) is trivially true.

In the case (ii) we can find two elements $j_v, j'_v \in J'_v$ such that $X_v^{j_v} = 1$, and $X_v^{j'_v} = 0$. If $K' = K \setminus \{j'_v\}$ then $(|K'| - |D^*|) = k$ and, since $j'_v \in C_1$, the connected components of L' are obtained by taking the connected components of $C'_1 = C_1 \setminus \{j'_v\}$ together with C_2, \dots, C_r . Hence, by the induction assumption

$$\text{pr}(Z_{L'} = z_{L'}) = \text{pr}(Z_{C'_1} = z_{C'_1}) \prod_{l=2}^r \text{pr}(Z_{C_l} = z_{C_l}) \quad (8)$$

where $\text{pr}(Z_{C'_1} = z_{C'_1})$ in (8) further factorizes according to the connected components of C'_1 . The result follows from (8) by noticing that $X_v^{j'_v} = 0$ is implied by $X_v^{j_v} = 1$ so that $\text{pr}(X_v^{j_v} = 1) = \text{pr}(X_v^{j_v} = 1, X_v^{j'_v} = 0)$ and, more generally, both $\text{pr}(Z_L = z_L) = \text{pr}(Z_{L'} = z_{L'})$ and $\text{pr}(Z_{C_1} = z_{C_1}) = \text{pr}(Z_{C'_1} = z_{C'_1})$.

In the case (iii) we can find two elements $j_v, j'_v \in J'_v$ such that $X_v^{j_v} = 0$ and $X_v^{j'_v} = 0$ and we can write both

$$\text{pr}(Z_L = z_L) = \text{pr}(Z_{L'} = z_{L'}) - \text{pr}(Z_{L'} = z_{L'}, X_v^{j'_v} = 1) \quad (9)$$

and

$$\text{pr}(Z_{C_1} = z_{C_1}) = \text{pr}(Z_{C'_1} = z_{C'_1}) - \text{pr}(Z_{C'_1} = z_{C'_1}, X_v^{j'_v} = 1), \quad (10)$$

where, as above, $L' = L \setminus \{j'_v\}$ and $C'_1 = C_1 \setminus \{j'_v\}$. By the induction assumption, $\text{pr}(Z_{L'} = z_{L'})$ in (9) factorizes as in (8). The probability $\text{pr}(Z_{L'} = z_{L'}, X_v^{j'_v} = 1)$ in (9) belongs to the case (ii) above because here $X_v^{j_v} = 0$ and $X_v^{j'_v} = 1$ so that

$$\text{pr}(Z_{L'} = z_{L'}, X_v^{j'_v} = 1) = \text{pr}(Z_{C'_1} = z_{C'_1}, X_v^{j'_v} = 1) \prod_{l=2}^r \text{pr}(Z_{C_l} = z_{C_l}). \quad (11)$$

Hence, by applying the factorizations (8) and (11) in (9) one obtains

$$\text{pr}(Z_L = z_L) = \{\text{pr}(Z_{C'_1} = z_{C'_1}) - \text{pr}(Z_{C'_1} = z_{C'_1}, X_v^{j'_v} = 1)\} \prod_{l=2}^r \text{pr}(Z_{C_l} = z_{C_l})$$

and then, by (10), $\text{pr}(Z_L = z_L) = \prod_{l=1}^r \text{pr}(Z_{C_l} = z_{C_l})$ as required.

Proof of Lemma 5.4

Let K be a primary subset of J_B . In this case, by Lemma 5.1, there exists a unique subset $D \subseteq B$ such that $K = j_D \in \mathcal{J}_D$ and, therefore, $X_K = X_{j_D} = (X_v^{j_v})_{j_v \in j_D}$. In order to show that $\text{pr}(X_K = 1_K, X_{J_B \setminus K} = 0_{J_B \setminus K}) = \text{pr}(Y_D = j_D, Y_{B \setminus D} = 0_{B \setminus D})$ we notice that (i) $J_B = J_{B \setminus D} \dot{\cup} J_D$ and therefore also $J_B \setminus j_D = J_{B \setminus D} \dot{\cup} (J_D \setminus j_D)$ because $j_D \subseteq J_D$; (ii) $X_{j_D} = 1_{j_D}$ implies that $X_{J_D \setminus j_D} = 0_{J_D \setminus j_D}$; (iii) it follows from (4) that both $X_{j_D} = 1_{j_D}$ iff $Y_D = j_D$, and $X_{J_{B \setminus D}} = 0_{J_{B \setminus D}}$ iff $Y_{B \setminus D} = 0_{B \setminus D}$. Hence, by applying (i) to (iii) one obtains

$$\begin{aligned} \text{pr}(X_K = 1_K, X_{J_B \setminus K} = 0_{J_B \setminus K}) &= \text{pr}(X_{j_D} = 1_{j_D}, X_{J_B \setminus j_D} = 0_{J_B \setminus j_D}) \\ &= \text{pr}(X_{j_D} = 1_{j_D}, X_{J_{B \setminus D}} = 0_{J_{B \setminus D}}, X_{J_D \setminus j_D} = 0_{J_D \setminus j_D}) \\ &= \text{pr}(X_{j_D} = 1_{j_D}, X_{J_{B \setminus D}} = 0_{J_{B \setminus D}}) \\ &= \text{pr}(Y_D = j_D, Y_{B \setminus D} = 0_{B \setminus D}), \end{aligned}$$

as required. The fact that for $j_D \neq \emptyset$ it holds that $\text{pr}(X_K = 1_K) = \text{pr}(Y_D = j_D)$ follows from (iii) above, and this completes the proof of the first statement. We prove the second statement by noticing that, if $K \subseteq J_B$ is non-primary, then, it follows from the definition of primary subset that there exists a $v \in B$ such that $|K \cap J_v| > 1$. Consequently, there exists a pair $j_v, j'_v \in J_v$ with $j_v \neq j'_v$ such that $j_v, j'_v \in K$. In this case, $\text{pr}(X_K = 1_K) \leq \text{pr}(X_v^{j_v} = 1, X_v^{j'_v} = 1) = \text{pr}(Y_v = j_v, Y_v = j'_v) = 0$. This also implies that $\text{pr}(X_K = 1_K, X_{J_D \setminus K} = 0_{J_D \setminus K}) = 0$ because $\text{pr}(X_K = 1_K, X_{J_D \setminus K} = 0_{J_D \setminus K}) \leq \text{pr}(X_K = 1_K) = 0$.

A basic lemma

This lemma will be used in the proof of Theorem 5.5 below.

Lemma A.1. *Let X_{J_B} be the binary expansion of Y_B . Furthermore, let $\tilde{\mu} = (\tilde{\mu}_K)_{K \subseteq J_B}$ and $\tilde{\gamma} = (\tilde{\gamma}_K)_{K \subseteq J_B}$ be the Möbius and the LML parameter of X_{J_B} , respectively, whereas μ and γ are the Möbius and the LML parameter of Y_V , respectively. If K is a primary subset of J_B then we can write $K = j_D$, by Lemma 5.1 and it holds that both $\mu^{j_D} = \tilde{\mu}_{j_D}$ and $\gamma^{j_D} = \tilde{\gamma}_{j_D}$.*

Furthermore, for every $K \subseteq J_B$ that is non-primary it holds that $\tilde{\mu}_K = 0$ and, consequently, $\tilde{\gamma}_K$ is not well-defined.

Proof. For every $K \subseteq J_B$ it holds by (1) that $\tilde{\mu}_K = \text{pr}(X_K = 1_K)$. Recall that, if K is a primary subset of J_B then, by Lemma 5.1, there exists a $D \subseteq B$ and a $j_D \in \mathcal{J}_D$ such that $K = j_D$. Hence, if $K = j_D$ is a primary subset of J_D then it follows from Lemma 5.4 and (5) that $\tilde{\mu}_{j_D} = \text{pr}(X_{j_D} = 1_{j_D}) = \text{pr}(Y_D = j_D) = \mu^{j_D}$. Conversely, if K is non-primary then $\tilde{\mu}_K = \text{pr}(X_K = 1_K) = 0$ by Lemma 5.4.

Similarly, by definition, $\tilde{\gamma}_K = \sum_{E \subseteq K} (-1)^{|K \setminus E|} \log \tilde{\mu}_E$ which is well-defined if and only if $\tilde{\mu}_K > 0$. Hence, if K is non-primary then the quantity $\tilde{\gamma}_K$ is not well-defined because in this case $\tilde{\mu}_K = 0$, as shown above. On the other hand, if $K = j_D$ is a primary subset of J_B then it is not difficult to check that (i) the set $\{E | E \subseteq j_D\}$ can be alternatively written in the form $\{j_H = (j_D)_H | H \subseteq D\}$, (ii) if $H \subseteq D$ then $(j_D)_H = j_H \in \mathcal{J}_H$ so that $\tilde{\mu}_{j_H} = \mu^{j_H}$, and (iii) for every $H \subseteq D$ it holds that $|j_D| = |D|$, $|j_H| = |H|$ and $j_H \subseteq j_D$ so that $|j_D \setminus j_H| = |D \setminus H|$. By using points (i) to (iii) above one obtains

$$\begin{aligned} \tilde{\gamma}_{j_D} &= \sum_{E \subseteq j_D} (-1)^{|j_D \setminus E|} \log \tilde{\mu}_E \\ &= \sum_{H \subseteq D} (-1)^{|j_D \setminus j_H|} \log \tilde{\mu}_{j_H} \\ &= \sum_{H \subseteq D} (-1)^{|D \setminus H|} \log \mu^{j_H} \\ &= \gamma^{j_D}. \end{aligned}$$

□

Proof of Theorem 5.5

We first consider the case where L is a primary subset of $P \cup J_B$, so that $L = Q \cup j_D$ as in Lemma 5.1. If we take the binary expansion of Y_V with respect to $J_V = J_P \dot{\cup} J_B$, that is equal to $X_{J_V} = (X_{J_P}, X_{J_B})$, then it follows from Lemma A.1 that

$$\mu^{j_{Q \cup D}} = \tilde{\mu}_{j_{Q \cup D}} \quad \text{and} \quad \gamma^{j_{Q \cup D}} = \tilde{\gamma}_{j_{Q \cup D}} \quad (12)$$

for every $j_Q \in \mathcal{J}_Q$. Recall that, since both the Möbius and the LML parameters satisfy the upward compatibility property then the computation of both $\tilde{\mu}_{j_{Q \cup D}}$ and $\tilde{\gamma}_{j_{Q \cup D}}$ can be carried out with respect to the distribution of (X_{j_Q}, X_{j_D}) that is a subvector of (X_{J_Q}, X_{j_D}) . It also follows from upward compatibility that, for every $j_Q \in \mathcal{J}_Q$, both $\tilde{\mu}_{j_D}^{j_Q}$ and $\tilde{\gamma}_{j_D}^{j_Q}$ can be computed with respect to the distribution of (Y_Q, X_{j_D}) . One should notice that X_{j_D} contains exactly one entry for every variable in Y_D and, as a consequence, the probability table of (Y_Q, X_{j_D}) is strictly positive because it can be obtained by collapsing some levels of the probability table of $Y_{Q \cup D}$, that is assumed to be strictly positive. For this reason it makes sense to consider (Y_Q, X_{j_D}) in place of Y_V in Lemma A.1 to show that

$$\tilde{\mu}_{j_D}^{j_Q} = \tilde{\mu}_{j_{Q \cup D}} \quad \text{and} \quad \tilde{\gamma}_{j_D}^{j_Q} = \tilde{\gamma}_{j_{Q \cup D}} \quad (13)$$

where the quantities $\tilde{\mu}_{j_Q \cup D}$ and $\tilde{\gamma}_{j_Q \cup D}$ in (13) coincide with same quantities in (12) because they are computed with respect to the binary expansion (X_{J_Q}, X_{j_D}) of (Y_Q, X_{j_D}) . Hence, (12) and (13) lead to $\mu^{j_Q \cup D} = \tilde{\mu}_{j_D}^{j_Q}$ and $\gamma^{j_Q \cup D} = \tilde{\gamma}_{j_D}^{j_Q}$ as required.

Consider now the case where L is non-primary. If we set $K = L \cap J_B$ and $Q = L \cap P$, then, for every $j_Q \in \mathcal{J}_Q$ it follows from definition of Möbius parameter and Lemma 5.4 that $\tilde{\mu}_{j_D}^{j_Q} = \text{pr}(Y_Q = j_Q, X_K = 1_K) \leq \text{pr}(X_K = 1_K) = 0$. This implies that $\tilde{\gamma}_{j_D}^{j_Q}$ is not well-defined because its computation involves the logarithm of $\tilde{\mu}_{j_D}^{j_Q} = 0$.

Proof of Theorem 5.2

We start this proof by remarking that, by Lemma 5.1, a subsets $L \subset P \cup J_B$ is such that $L \subset P \cup j_B$ for some $j_B \in \mathcal{J}_B$ if and only if it is a primary subset of $P \cup J_B$. This also implies that if $L \subset P \cup j_B$ we can write $L = Q \cup j_D$, as in Lemma 5.1.

By Lemma 5.3, the distribution of $Y^B = (Y_P, X_{J_B})$ satisfies the connected set Markov property with respect to \mathcal{G}^B if and only if the distribution of (Y_P, X_{j_B}) satisfies the connected set Markov property with respect to $\mathcal{G}_{P \cup j_B}^B$ for every $j_B \in \mathcal{J}_B$.

For every $j_B \in \mathcal{J}_B$, the random vector X_{j_B} contains exactly one entry for every variable in Y_B and, as a consequence, the probability table of (Y_P, X_{j_B}) is strictly positive because it can be obtained by collapsing some levels of the probability table of Y_V , that is strictly positive by assumption. Hence, we can apply Corollary 3.2, and it follows that the distribution of (Y_P, X_{j_B}) satisfies the connected set Markov property with respect to $\mathcal{G}_{P \cup j_B}^B$ if and only if for every set $L \subseteq P \cup j_B$ that is disconnected in $\mathcal{G}_{P \cup j_B}^B$ it holds that $\tilde{\gamma}_{j_D}^{j_Q} = 0$ for every $j_Q \in \mathcal{J}_Q$. Note that it is possible that either $Q = \emptyset$ or $D = \emptyset$. We can thus state that:

The distribution of $Y^B = (Y_P, X_{J_B})$ satisfies the connected set Markov property with respect to \mathcal{G}^B if and only if for every $j_B \in \mathcal{J}_B$ and every set $L \subseteq P \cup j_B$ that is disconnected in $\mathcal{G}_{P \cup j_B}^B$ it holds that $\tilde{\gamma}_{j_D}^{j_Q} = 0$ for every $j_Q \in \mathcal{J}_Q$.

Notice that: (a) a subsets $L \subseteq P \cup J_B$ is such that $L \subseteq P \cup j_B$ for some $j_B \in \mathcal{J}_B$ if and only if it is a primary subset of $P \cup J_B$, (b) $L \subset P \cup J_B$ is disconnected in $\mathcal{G}_{P \cup j_B}^B$ if and only if it is disconnected in \mathcal{G}^B and (c) $\tilde{\gamma}_{j_D}^{j_Q} = \gamma^{j_Q \cup D}$ by Corollary 5.5. Hence, by using (a), (b) and (c) the statement above can be rephrased as:

The distribution of $Y^B = (Y_P, X_{J_B})$ satisfies the connected set Markov property with respect to \mathcal{G}^B if and only if for every set $L \subseteq P \cup J_B$ that is a primary subset of $P \cup J_B$ and disconnected in \mathcal{G}^B it holds that $\gamma^{j_Q \cup D} = 0$ for every $j_Q \in \mathcal{J}_Q$,

and this completes the proof.

References

Balding, D. (2006). A tutorial on statistical methods for population association studies. *Nature Reviews Genetics* 7(10), 781–791.

- Banerjee, M. and T. Richardson (2003). On a dualization of graphical gaussian models: A correction note. *Scandinavian Journal of Statistics* 30(4), 817–820.
- Bergsma, W., M. Croon, and J. Hagenaars (2009). *Marginal models for dependent, clustered, and longitudinal categorical data*. London, UK: Springer.
- Colombo, D., M. Kalisch, P. Bühlmann, and R. T. S. (2012). Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics* 40(1), 294–321.
- Cox, D. R. and N. Wermuth (1993). Linear dependencies represented by chain graphs. *Statistical Science* 8(3), 204–218.
- Cox, D. R. and N. Wermuth (1996). *Multivariate dependencies. Models, analysis and interpretation*. London: Chapman and Hall.
- Drton, M. (2009). Discrete chain graph models. *Bernoulli* 15(3), 736–753.
- Drton, M. and T. Richardson (2008). Binary models for marginal independence. *Journal of the Royal Statistical Society: Series B* 70(2), 287–309.
- Ekholm, A., J. W. McDonald, and P. W. F. Smith (2000). Association models for a multivariate binary response. *Biometrics* 56(3), 712–718.
- Evans, R. J. and T. S. Richardson (2012). Marginal log-linear parameters for graphical Markov models. Technical report, arXiv:1105.6075v1[stat.ME].
- Glonek, G. J. N. and P. McCullagh (1995). Multivariate logistic models. *Journal of the Royal Statistical Society, Series B* 57(3), 533–546.
- Hirschhorn, J. and M. Daly (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 6(2), 95–108.
- Højsgaard, S. and S. Lauritzen (2008). Graphical gaussian models with edge and vertex symmetries. *Journal of the Royal Statistical Society: Series B* 70(5), 1005–1027.
- Kauermann, G. (1996). On a dualization of graphical Gaussian models. *Scandinavian Journal of Statistics* 23, 105–116.
- Lang, J. B. (1996). Maximum likelihood methods for a generalized class of log-linear models. *Annals of Statistics* 24(2), 726–752.
- Lauritzen, S. L. (1996). *Graphical Models*. Oxford, UK: Clarendon Press.
- Lupparelli, M., G. M. Marchetti, and W. P. Bergsma (2009). Parameterizations and fitting of bi-directed graph models to categorical data. *Scandinavian Journal of Statistics* 36(3), 559–576.

- Maathuis, M., M. Kalisch, and P. Bühlmann (2009). Estimating high-dimensional intervention effects from observational data. *The Annals of Statistics* 37(6A), 3133–3164.
- Pearl, J. and N. Wermuth (1994). When can association graphs admit a causal interpretation? In *Models and Data, Artificial Intelligence and Statistics IV* (P. Cheseman and W. Oldford ed.), pp. 205–214. New York: Springer-Verlag.
- Richardson, T. (2009). A factorization criterion for acyclic directed mixed graphs. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pp. 462–470. AUAI Press, Arlington (VA).
- Richardson, T. S. (2003). Markov property for acyclic directed mixed graphs. *Scandinavian Journal of Statistics* 30(1), 145–157.
- Roverato, A., M. Lupparelli, and L. La Rocca (2013). Log-mean linear models for binary data. *Biometrika*, in press.
- Rudas, T., W. Bergsma, and R. Nemeth (2010). Marginal log-linear parameterization of conditional independence models. *Biometrika* 97(4), 1006–1012.
- Wermuth, N. and K. Sadeghi (2012). Sequences of regressions and their independences. *TEST* 21(2), 215–252.